

# SEQUENCE LISTING

<110> Koffas, Mattheos  
Odom, J. Martin  
Schenzle, Andreas J.  
Norton, Kelley C.  
Tomb, Jean-Francois  
Rouviere, Pierre  
Picataggio, Stephen  
Cheng, Qiong

<120> Genes Involved in Isoprenoid Compounds Production

<130> CL1646 US NA

<140>

<141>

<150> 60/229,907

<151> September 1, 2001

<160> 24

<170> Microsoft Office 97

<210> 1

<211> 1860

<212> DNA

<213> Methylobionas 16a

<220>

<223> ORF1

<400> 1

atgaaactga	ccaccgacta	tcccttgctt	aaaaacatcc	acacgccggc	ggacatacgc	60
gcgctgtcca	aggaccagct	ccagcaactg	gctgacgagg	tgccgggcta	tctgaccac	120
acggtcagca	tttccggcgg	ccatttttgcg	gccggcctcg	gcaccgtgga	actgaccgtg	180
gccttgcat	atgtgttcaa	tacccccgctc	gatcagttgg	tctgggacgt	gggccatcag	240
gcctatccgc	acaagattct	gaccggctcg	aaggagcgca	tgccgaccat	tcgcaccctg	300
ggcgggggtg	cagcctttcc	ggcgcgggac	gagagcgaat	acgatgcctt	cggcgctcggc	360
cattccagca	cctcgatcag	cgcggcactg	ggcatggcca	ttgcgtcgca	gctgcgcggc	420
gaagacaaga	agatggtagc	catcatcggc	gacggttcca	tcaccggcgg	catggcctat	480
gaggcgatga	atcatgccgg	cgatgtgaat	gccaacctgc	tggtgatctt	gaacgacaac	540
gatatgtcga	tctcgccgcc	ggtcggggcg	atgaacaatt	atctgaccaa	ggtgttgtcg	600
agcaagtttt	attcgctcgt	gcgggaagag	agcaagaaag	ctctggccaa	gatgccgtcg	660
gtgtgggaac	tgccgcgcaa	gaccgaggaa	cacgtgaagg	gcattgatcgt	gcccgggtacc	720
ttgttcgagg	aattgggctt	caattatttc	ggcccgatcg	acggccatga	tgctcgagatg	780
ctgggtgtcga	ccctggaaaa	tctgaaggat	ttgaccgggc	cggatttcct	gcattgtggtg	840
accaagaagg	gcaaaggcta	tgccgccagcc	gagaaagacc	cgttggccta	ccatggcgtg	900
ccggcttttcg	atccgaccaa	ggatttcctg	cccaaggcgg	cgccgtcgcc	gcattccgacc	960
tataccgagg	tggtcggccg	ctggctgtgc	gacatggcgg	ctcaagacga	gcgcttgctg	1020
ggcatcacgc	cggcgatgcg	cgaaggctct	ggtttggtgg	aattctcaca	gaaatttccg	1080
aatcgctatt	tcgatgtcgc	catcgccgag	cagcatgcgg	tgaccttggc	cgcgggccag	1140
gcttgccagg	gcgccaagcc	ggtggtggcg	atttattcca	ccttcctgca	acgcgggttac	1200
gatcagttga	tccacgacgt	ggccttgacg	aacttagata	tgctctttgc	actggatcgt	1260
gccggcttgg	tcggcccggg	tggaaccgacc	catgctggcg	cctttgatta	cagctacatg	1320
cgctgtattc	cgaacatgct	gatcatggct	ccagccgacg	agaacgagtg	caggcagatg	1380
ctgaccaccg	gcttccaaca	ccatggcccg	gcttcggtgc	gctatccgcg	cggcaaaagg	1440

```

ccccggggcg caatcgatcc gaccctgacc gcgctggaga tcggcaaggc cgaagtcaga 1500
caccacggca gccgcacgc cattctggcc tggggcagca tggtcacgcc tgccgtcgaa 1560
gccggcaagc agctggggcg gacggtggtg aacatgcgtt tcgtcaagcc gttcgatcaa 1620
gccttgggtgc tggaattggc caggacgcac gatgtgttcg tcaccgtcga ggaaaacgtc 1680
atcgccggcg gcgctggcag tgcgatcaac accttcctgc aggcgcagaa ggtgctgatg 1740
ccggtctgca acatcggcct gcccgaccgc ttcgtcgagc aaggtagtcg cgaggaattg 1800
ctcagcctgg tcggcctcga cagcaagggc atcctcgcca ccatcgaaca gttttgcgct 1860

```

<210> 2

<211> 620

<212> PRT

<213> Methylomonas 16a

<220>

<223> Amino acid sequences encoded by ORF1

<400> 2

```

Met Lys Leu Thr Thr Asp Tyr Pro Leu Leu Lys Asn Ile His Thr Pro
  1                      5                      10                      15

Ala Asp Ile Arg Ala Leu Ser Lys Asp Gln Leu Gln Gln Leu Ala Asp
          20                      25                      30

Glu Val Arg Gly Tyr Leu Thr His Thr Val Ser Ile Ser Gly Gly His
          35                      40                      45

Phe Ala Ala Gly Leu Gly Thr Val Glu Leu Thr Val Ala Leu His Tyr
          50                      55                      60

Val Phe Asn Thr Pro Val Asp Gln Leu Val Trp Asp Val Gly His Gln
          65                      70                      75                      80

Ala Tyr Pro His Lys Ile Leu Thr Gly Arg Lys Glu Arg Met Pro Thr
          85                      90                      95

Ile Arg Thr Leu Gly Gly Val Ser Ala Phe Pro Ala Arg Asp Glu Ser
          100                     105                     110

Glu Tyr Asp Ala Phe Gly Val Gly His Ser Ser Thr Ser Ile Ser Ala
          115                     120                     125

Ala Leu Gly Met Ala Ile Ala Ser Gln Leu Arg Gly Glu Asp Lys Lys
          130                     135                     140

Met Val Ala Ile Ile Gly Asp Gly Ser Ile Thr Gly Gly Met Ala Tyr
          145                     150                     155                     160

Glu Ala Met Asn His Ala Gly Asp Val Asn Ala Asn Leu Leu Val Ile
          165                     170                     175

Leu Asn Asp Asn Asp Met Ser Ile Ser Pro Pro Val Gly Ala Met Asn
          180                     185                     190

Asn Tyr Leu Thr Lys Val Leu Ser Ser Lys Phe Tyr Ser Ser Val Arg
          195                     200                     205

Glu Glu Ser Lys Lys Ala Leu Ala Lys Met Pro Ser Val Trp Glu Leu
          210                     215                     220

```

Ala	Arg	Lys	Thr	Glu	Glu	His	Val	Lys	Gly	Met	Ile	Val	Pro	Gly	Thr	
225					230					235					240	
Leu	Phe	Glu	Glu	Leu	Gly	Phe	Asn	Tyr	Phe	Gly	Pro	Ile	Asp	Gly	His	
				245					250					255		
Asp	Val	Glu	Met	Leu	Val	Ser	Thr	Leu	Glu	Asn	Leu	Lys	Asp	Leu	Thr	
			260					265					270			
Gly	Pro	Val	Phe	Leu	His	Val	Val	Thr	Lys	Lys	Gly	Lys	Gly	Tyr	Ala	
		275					280					285				
Pro	Ala	Glu	Lys	Asp	Pro	Leu	Ala	Tyr	His	Gly	Val	Pro	Ala	Phe	Asp	
	290					295					300					
Pro	Thr	Lys	Asp	Phe	Leu	Pro	Lys	Ala	Ala	Pro	Ser	Pro	His	Pro	Thr	
305					310					315					320	
Tyr	Thr	Glu	Val	Phe	Gly	Arg	Trp	Leu	Cys	Asp	Met	Ala	Ala	Gln	Asp	
				325					330					335		
Glu	Arg	Leu	Leu	Gly	Ile	Thr	Pro	Ala	Met	Arg	Glu	Gly	Ser	Gly	Leu	
			340					345					350			
Val	Glu	Phe	Ser	Gln	Lys	Phe	Pro	Asn	Arg	Tyr	Phe	Asp	Val	Ala	Ile	
		355					360					365				
Ala	Glu	Gln	His	Ala	Val	Thr	Leu	Ala	Ala	Gly	Gln	Ala	Cys	Gln	Gly	
		370				375					380					
Ala	Lys	Pro	Val	Val	Ala	Ile	Tyr	Ser	Thr	Phe	Leu	Gln	Arg	Gly	Tyr	
385					390					395					400	
Asp	Gln	Leu	Ile	His	Asp	Val	Ala	Leu	Gln	Asn	Leu	Asp	Met	Leu	Phe	
				405					410					415		
Ala	Leu	Asp	Arg	Ala	Gly	Leu	Val	Gly	Pro	Asp	Gly	Pro	Thr	His	Ala	
			420					425					430			
Gly	Ala	Phe	Asp	Tyr	Ser	Tyr	Met	Arg	Cys	Ile	Pro	Asn	Met	Leu	Ile	
		435					440					445				
Met	Ala	Pro	Ala	Asp	Glu	Asn	Glu	Cys	Arg	Gln	Met	Leu	Thr	Thr	Gly	
		450				455					460					
Phe	Gln	His	His	Gly	Pro	Ala	Ser	Val	Arg	Tyr	Pro	Arg	Gly	Lys	Gly	
465					470					475					480	
Pro	Gly	Ala	Ala	Ile	Asp	Pro	Thr	Leu	Thr	Ala	Leu	Glu	Ile	Gly	Lys	
				485					490					495		
Ala	Glu	Val	Arg	His	His	Gly	Ser	Arg	Ile	Ala	Ile	Leu	Ala	Trp	Gly	
			500					505					510			
Ser	Met	Val	Thr	Pro	Ala	Val	Glu	Ala	Gly	Lys	Gln	Leu	Gly	Ala	Thr	
		515					520					525				
Val	Val	Asn	Met	Arg	Phe	Val	Lys	Pro	Phe	Asp	Gln	Ala	Leu	Val	Leu	
		530				535					540					

Glu Leu Ala Arg Thr His Asp Val Phe Val Thr Val Glu Glu Asn Val  
 545 550 555 560  
 Ile Ala Gly Gly Ala Gly Ser Ala Ile Asn Thr Phe Leu Gln Ala Gln  
 565 570 575  
 Lys Val Leu Met Pro Val Cys Asn Ile Gly Leu Pro Asp Arg Phe Val  
 580 585 590  
 Glu Gln Gly Ser Arg Glu Glu Leu Leu Ser Leu Val Gly Leu Asp Ser  
 595 600 605  
 Lys Gly Ile Leu Ala Thr Ile Glu Gln Phe Cys Ala  
 610 615 620

<210> 3  
 <211> 1182  
 <212> DNA  
 <213> *Methylomonas* 16a

<220>  
 <223> ORF2

<400> 3  
 atgaaaggta tttgcatatt gggcgctacc ggttcgatcg gtgtcagcac gctggatgtc 60  
 gttgccaggc atccggataa atatcaagtc gttgcgctga cgcgcaacgg caatatcgac 120  
 gcattgtatg aacaatgcct ggcccacccat ccggagtatg cgggtggtggt catggaaagc 180  
 aaggtagcag agttcaaaca gcgcattgcc gcttcgccgg tagcggatat caaggtcttg 240  
 tcgggtagcg aggccttgca acaggtggcc acgctggaaa acgtcgatac ggtgatggcg 300  
 gctatcgctc gcgcggccgg attgttgccg accttggccg cggccaaggc cggcaaaacc 360  
 gtgctgttgg ccaacaagga agccttggtg atgtcgggac aaatcttcat gcaggccgctc 420  
 agcgattccg gcgctgtgtt gctgccgata gacagcgagc acaacgccat ctttcagtgc 480  
 atgccggcgg gttatacgcc aggccataca gccaaacagg cgcgcgcgat tttattgacc 540  
 gcttcgggtg gcccatctcg acggacgccg atagaaacgt tgtccagcgt cacgccggat 600  
 caggccggtg cccatcctaa atgggacatg gggcgcaaga tttcggtcga ttccgccacc 660  
 atgatgaaca aaggtctcga actgatcgaa gcctgcttgt tgttcaacat ggagcccgcg 720  
 cagattgaag tcgtcattca tccgcagagc atcattcatt cgatggtgga ctatgtcgat 780  
 ggttcgggtt tggcgcagat gggtaatccc gacatgcgca cgccgatagc gcacgcgatg 840  
 gcctggccgg aacgctttga ctctggtgtg gcgccgctgg atattttcga agtagggcac 900  
 atggatttcg aaaaaccoga cttgaaacgg tttccttgtc tgagattggc ttatgaagcc 960  
 atcaagtctg gtggaattat gccaaacggt tgaacgcag ccaatgaaat tgctgtcgaa 1020  
 gcgtttttta atgaagaagt caaattcact gacatcgcgg tcatcatcga gcgcagcatg 1080  
 gccagttta aaccggacga tgccggcagc ctcgaattgg ttttcaggc cgatcaagat 1140  
 gcgcgcgagg tggctagaga catcatcaag accttggtag ct 1182

<210> 4  
 <211> 394  
 <212> PRT  
 <213> *Methylomonas* 16a

<220>  
 <223> Amino acid sequences encoded by ORF2

<400> 4  
 Met Lys Gly Ile Cys Ile Leu Gly Ala Thr Gly Ser Ile Gly Val Ser  
 1 5 10 15

Thr	Leu	Asp	Val	Val	Ala	Arg	His	Pro	Asp	Lys	Tyr	Gln	Val	Val	Ala	
			20					25					30			
Leu	Thr	Ala	Asn	Gly	Asn	Ile	Asp	Ala	Leu	Tyr	Glu	Gln	Cys	Leu	Ala	
		35					40					45				
His	His	Pro	Glu	Tyr	Ala	Val	Val	Val	Met	Glu	Ser	Lys	Val	Ala	Glu	
	50					55					60					
Phe	Lys	Gln	Arg	Ile	Ala	Ala	Ser	Pro	Val	Ala	Asp	Ile	Lys	Val	Leu	
65					70					75					80	
Ser	Gly	Ser	Glu	Ala	Leu	Gln	Gln	Val	Ala	Thr	Leu	Glu	Asn	Val	Asp	
				85					90					95		
Thr	Val	Met	Ala	Ala	Ile	Val	Gly	Ala	Ala	Gly	Leu	Leu	Pro	Thr	Leu	
			100					105					110			
Ala	Ala	Ala	Lys	Ala	Gly	Lys	Thr	Val	Leu	Leu	Ala	Asn	Lys	Glu	Ala	
		115					120					125				
Leu	Val	Met	Ser	Gly	Gln	Ile	Phe	Met	Gln	Ala	Val	Ser	Asp	Ser	Gly	
	130					135					140					
Ala	Val	Leu	Leu	Pro	Ile	Asp	Ser	Glu	His	Asn	Ala	Ile	Phe	Gln	Cys	
145					150					155					160	
Met	Pro	Ala	Gly	Tyr	Thr	Pro	Gly	His	Thr	Ala	Lys	Gln	Ala	Arg	Arg	
				165					170					175		
Ile	Leu	Leu	Thr	Ala	Ser	Gly	Gly	Pro	Phe	Arg	Arg	Thr	Pro	Ile	Glu	
			180					185					190			
Thr	Leu	Ser	Ser	Val	Thr	Pro	Asp	Gln	Ala	Val	Ala	His	Pro	Lys	Trp	
		195					200					205				
Asp	Met	Gly	Arg	Lys	Ile	Ser	Val	Asp	Ser	Ala	Thr	Met	Met	Asn	Lys	
	210					215					220					
Gly	Leu	Glu	Leu	Ile	Glu	Ala	Cys	Leu	Leu	Phe	Asn	Met	Glu	Pro	Asp	
225					230					235					240	
Gln	Ile	Glu	Val	Val	Ile	His	Pro	Gln	Ser	Ile	Ile	His	Ser	Met	Val	
				245					250					255		
Asp	Tyr	Val	Asp	Gly	Ser	Val	Leu	Ala	Gln	Met	Gly	Asn	Pro	Asp	Met	
			260					265					270			
Arg	Thr	Pro	Ile	Ala	His	Ala	Met	Ala	Trp	Pro	Glu	Arg	Phe	Asp	Ser	
		275					280					285				
Gly	Val	Ala	Pro	Leu	Asp	Ile	Phe	Glu	Val	Gly	His	Met	Asp	Phe	Glu	
	290					295					300					
Lys	Pro	Asp	Leu	Lys	Arg	Phe	Pro	Cys	Leu	Arg	Leu	Ala	Tyr	Glu	Ala	
305					310					315					320	
Ile	Lys	Ser	Gly	Gly	Ile	Met	Pro	Thr	Val	Leu	Asn	Ala	Ala	Asn	Glu	
				325					330					335		

Ile Ala Val Glu Ala Phe Leu Asn Glu Glu Val Lys Phe Thr Asp Ile  
340 345 350

Ala Val Ile Ile Glu Arg Ser Met Ala Gln Phe Lys Pro Asp Asp Ala  
355 360 365

Gly Ser Leu Glu Leu Val Leu Gln Ala Asp Gln Asp Ala Arg Glu Val  
370 375 380

Ala Arg Asp Ile Ile Lys Thr Leu Val Ala  
385 390

<210> 5  
<211> 693  
<212> DNA  
<213> Methylomonas 16a

<220>  
<223> ORF3

<400> 5  
atgaacccaa ccatccaatg ctggggccgctc gtgcccgcag ccggcgctcgg caaacgcatg 60  
caagccgata gcccacaaca atattttaccg cttgccggta aaacgggtcat cgaacacaca 120  
ctgactcgac tacttgagtc cgacgccttc caaaaagttg cgggtggcgat ttccgctcgaa 180  
gacccttatt ggccctgaact gtccatagcc aaacaccccg acatcatcac cgcgccctggc 240  
ggcaagggaac gcgccgactc ggtgctgtct gcactgaagg ctttagaaga tatagccagc 300  
gaaaatgatt ggggtgctggt acacgacgcc gcccgcccct gcttgacggg cagcgacatc 360  
caccttcaaa tcgatacctt aaaaaatgac ccggctcggcg gcacccctggc cttgagttcg 420  
cacgacacat tgaaacacgt ggatgggtgac acgatcaccg caaccataga cagaaagcac 480  
gtctggcgcg ccttgacgcc gcaaattgtt aaatacggca tgttgcgaga cgcgttgcaa 540  
cgaaccgaag gcaatccggc cgtcaccgac gaagccagtg cgctggaact tttgggcat 600  
aaacccaaaa tcgtggaagg ccgcccggac aacatcaaaa tcacccgccc ggaagatttg 660  
gccctggcac aattttatat ggagcaacaa gca 693

<210> 6  
<211> 231  
<212> PRT  
<213> Methylomonas 16a

<220>  
<223> Amino acid sequences encoded by ORF3

<400> 6  
Met Asn Pro Thr Ile Gln Cys Trp Ala Val Val Pro Ala Ala Gly Val  
1 5 10 15

Gly Lys Arg Met Gln Ala Asp Arg Pro Lys Gln Tyr Leu Pro Leu Ala  
20 25 30

Gly Lys Thr Val Ile Glu His Thr Leu Thr Arg Leu Leu Glu Ser Asp  
35 40 45

Ala Phe Gln Lys Val Ala Val Ala Ile Ser Val Glu Asp Pro Tyr Trp  
50 55 60

Pro Glu Leu Ser Ile Ala Lys His Pro Asp Ile Ile Thr Ala Pro Gly  
65 70 75 80



<212> PRT

<213> Methylomonas 16a

<220>

<223> Amino acid sequences encoded by ORF4

<400> 8

```
Met Asp Tyr Ala Ala Gly Trp Gly Glu Arg Trp Pro Ala Pro Ala Lys
 1          5          10          15

Leu Asn Leu Met Leu Arg Ile Thr Gly Arg Arg Pro Asp Gly Tyr His
          20          25          30

Leu Leu Gln Thr Val Phe Gln Met Leu Asp Leu Cys Asp Trp Leu Thr
          35          40          45

Phe His Pro Val Asp Asp Gly Arg Val Thr Leu Arg Asn Pro Ile Ser
          50          55          60

Gly Val Pro Glu Gln Asp Asp Leu Thr Val Arg Ala Ala Asn Leu Leu
          65          70          75          80

Lys Ser His Thr Gly Cys Val Arg Gly Val Cys Ile Asp Ile Glu Lys
          85          90          95

Asn Leu Pro Met Gly Gly Gly Leu Gly Gly Gly Ser Ser Asp Ala Ala
          100          105          110

Thr Thr Leu Val Val Leu Asn Arg Leu Trp Gly Leu Gly Leu Ser Lys
          115          120          125

Arg Glu Leu Met Asp Leu Gly Leu Arg Leu Gly Ala Asp Val Pro Val
          130          135          140

Phe Val Phe Gly Cys Ser Ala Trp Gly Glu Gly Val Ser Glu Asp Leu
          145          150          155          160

Gln Ala Ile Thr Leu Pro Glu Gln Trp Phe Val Ile Ile Lys Pro Asp
          165          170          175

Cys His Val Asn Thr Gly Glu Ile Phe Ser Ala Glu Asn Leu Thr Arg
          180          185          190

Asn Ser Ala Val Val Thr Met Ser Asp Phe Leu Ala Gly Asp Asn Arg
          195          200          205

Asn Asp Cys Ser Glu Val Val Cys Lys Leu Tyr Arg Pro Val Lys Asp
          210          215          220

Ala Ile Asp Ala Leu Leu Cys Tyr Ala Glu Ala Arg Leu Thr Gly Thr
          225          230          235          240

Gly Ala Cys Val Phe Ala Gln Phe Cys Asn Lys Glu Asp Ala Glu Ser
          245          250          255

Ala Leu Glu Gly Leu Lys Asp Arg Trp Leu Val Phe Leu Ala Lys Gly
          260          265          270

Leu Asn Gln Ser Ala Leu Tyr Lys Lys Leu Glu Gln Gly
          275          280          285
```



<210> 9  
 <211> 471  
 <212> DNA  
 <213> Methylobionas 16a

<220>  
 <223> ORF5

<400> 9  
 atgatacgcg taggcatggg ttacgacgtg caccgtttca acgacggcga ccacatcatt 60  
 ttgggcggcg tcaaaatccc ttatgaaaaa ggcctggaag cccattccga cggcgacgtg 120  
 gtgctgcacg cattggccga cgccatcttg ggagccgccc ctttgggcca catcgccaaa 180  
 catttcccgg acaccgaccc caatttcaag ggcgcccaga gcagggtgct actgcgccac 240  
 gtgtacggca tcgtcaagga aaaaggctat aaactgggtca acgccgacgt gaccatcatc 300  
 gctcaggcgc cgaagatgct gccacacgtg cccggcatgc gcgccaacat tgccgccgat 360  
 ctggaaaccg atgtcgattt cattaatgta aaagccacga cgaccgagaa actgggcttt 420  
 gagggccgta aggaaggcat cgccgtgcag gctgtggtgt tgatagaacg c 471

<210> 10  
 <211> 157  
 <212> PRT  
 <213> Methylobionas 16a

<220>  
 <223> Amino acid sequences encoded by ORF5

<400> 10  
 Met Ile Arg Val Gly Met Gly Tyr Asp Val His Arg Phe Asn Asp Gly  
 1 5 10 15  
 Asp His Ile Ile Leu Gly Gly Val Lys Ile Pro Tyr Glu Lys Gly Leu  
 20 25 30  
 Glu Ala His Ser Asp Gly Asp Val Val Leu His Ala Leu Ala Asp Ala  
 35 40 45  
 Ile Leu Gly Ala Ala Ala Leu Gly Asp Ile Gly Lys His Phe Pro Asp  
 50 55 60  
 Thr Asp Pro Asn Phe Lys Gly Ala Asp Ser Arg Val Leu Leu Arg His  
 65 70 75 80  
 Val Tyr Gly Ile Val Lys Glu Lys Gly Tyr Lys Leu Val Asn Ala Asp  
 85 90 95  
 Val Thr Ile Ile Ala Gln Ala Pro Lys Met Leu Pro His Val Pro Gly  
 100 105 110  
 Met Arg Ala Asn Ile Ala Ala Asp Leu Glu Thr Asp Val Asp Phe Ile  
 115 120 125  
 Asn Val Lys Ala Thr Thr Thr Glu Lys Leu Gly Phe Glu Gly Arg Lys  
 130 135 140  
 Glu Gly Ile Ala Val Gln Ala Val Val Leu Ile Glu Arg  
 145 150 155

<210> 11  
 <211> 1632  
 <212> DNA  
 <213> *Methylomonas* 16a

<220>  
 <223> ORF6

<400> 11  
 atgacaaaat tcatctttat caccggcggc gtggtgtcat ccttgggaaa agggatagcc 60  
 gcctcctccc tggcgcgat tctggaagac cgcggcctca aagtcactat cacaaaactc 120  
 gatccctaca tcaacgtcga ccccggcacc atgagcccgt ttcaacacgg cgaggtgttc 180  
 gtgaccgaag acggtgccga aaccgatttg gaccttggcc attacgaacg gtttttgaaa 240  
 accacgatga ccaagaaaaa caacttcacc accggtcagg ttacgagca ggtattacgc 300  
 aacgagcgca aaggtgatta tcttggcgcg accgtgcaag tcattccaca tatcaccgac 360  
 gaaatcaaac gccgggtgta tgaaagcgcc gaagggaag atgtggcatt gatcgaagtc 420  
 ggcggcacgg tggcgacat cgaatcgta ccgtttctgg aaaccatacg ccagatgggc 480  
 gtggaactgg gtcgtagccg cgccttggtc attcatttga cgctggtgcc ttacatcaaa 540  
 tcggccggcg aactgaaaac caagcccacc cagcattcgg tcaaagaact gcgcaccatc 600  
 gggattcagc cggacatttt gatctgtcgt tcagaacaac cgatcccggc cagtgaacgc 660  
 cgcaagatcg cgctattttac caatgtcgcc gaaaaggcgg tgatttccgc gatcgatgcc 720  
 gacaccattt accgcatttc gctattgctg cgcgacaag gcctggacga cctgggtggtc 780  
 gatcagttgc gcctggacgt accagcggcg gatttatcgg cctgggaaaa ggtcgtcgat 840  
 ggcctgactc atccgaccga cgaagtcagc attgcatcg tcggtaaata tgtcgaccac 900  
 accgatgcct acaaatcgct gaatgaagcc ctgattcatg ccggcattca cgcgcgccac 960  
 aaggtgcaaa tcagctacat cgactccgaa accatagaag ccgaaggcac cgccaaattg 1020  
 aaaaacgtcg atgcgatcct ggtgccgggt ggtttcggcg aacgcggcgt ggaaggcaag 1080  
 atttctaccg tgcgttttgc ccgcgagaac aaaatcccgt atttgggcat ttgcttgggc 1140  
 atgcaatcgg cggtaatcga attcgcgcgc aacgtggttg gcctggaagg cgcgcacagc 1200  
 accgaattcc tgccgaaatc gccacaccct gtgatcggct tgatcaccga atggatggac 1260  
 gaagccggcg aactggtcac acgcgacgaa gattccgatc tggcgggcac gatgcgtctg 1320  
 ggcgcgcaaa aatgccgcct gaaggctgat tccttggctt ttcagttgta tcaaaaagac 1380  
 gtcatcaccg agcgtcaccg ccaccgctac gaattcaaca atcaatattt aaaacaactg 1440  
 gaagcggccg gcatgaaatt ttccggtaaa tcgctggacg gccgcctggt ggagatcatc 1500  
 gagctaccgg aacacccctg gttcctggcc tgccagttcc atcccgaatt cacctcgacg 1560  
 ccgcgtaacg gccacgcctt attttcgggc ttcgtcgaag cggccgcaa acacaaaaca 1620  
 caaggcacag ca 1632

<210> 12  
 <211> 544  
 <212> PRT  
 <213> *Methylomonas* 16a

<220>  
 <223> Amino acid sequences encoded by ORF6

<400> 12  
 Met Thr Lys Phe Ile Phe Ile Thr Gly Gly Val Val Ser Ser Leu Gly  
 1 5 10 15  
 Lys Gly Ile Ala Ala Ser Ser Leu Ala Ala Ile Leu Glu Asp Arg Gly  
 20 25 30  
 Leu Lys Val Thr Ile Thr Lys Leu Asp Pro Tyr Ile Asn Val Asp Pro  
 35 40 45  
 Gly Thr Met Ser Pro Phe Gln His Gly Glu Val Phe Val Thr Glu Asp  
 50 55 60

Gly	Ala	Glu	Thr	Asp	Leu	Asp	Leu	Gly	His	Tyr	Glu	Arg	Phe	Leu	Lys	
65					70					75					80	
Thr	Thr	Met	Thr	Lys	Lys	Asn	Asn	Phe	Thr	Thr	Gly	Gln	Val	Tyr	Glu	
				85					90					95		
Gln	Val	Leu	Arg	Asn	Glu	Arg	Lys	Gly	Asp	Tyr	Leu	Gly	Ala	Thr	Val	
			100					105					110			
Gln	Val	Ile	Pro	His	Ile	Thr	Asp	Glu	Ile	Lys	Arg	Arg	Val	Tyr	Glu	
		115					120					125				
Ser	Ala	Glu	Gly	Lys	Asp	Val	Ala	Leu	Ile	Glu	Val	Gly	Gly	Thr	Val	
	130					135					140					
Gly	Asp	Ile	Glu	Ser	Leu	Pro	Phe	Leu	Glu	Thr	Ile	Arg	Gln	Met	Gly	
145					150					155					160	
Val	Glu	Leu	Gly	Arg	Asp	Arg	Ala	Leu	Phe	Ile	His	Leu	Thr	Leu	Val	
				165					170					175		
Pro	Tyr	Ile	Lys	Ser	Ala	Gly	Glu	Leu	Lys	Thr	Lys	Pro	Thr	Gln	His	
			180					185					190			
Ser	Val	Lys	Glu	Leu	Arg	Thr	Ile	Gly	Ile	Gln	Pro	Asp	Ile	Leu	Ile	
		195					200					205				
Cys	Arg	Ser	Glu	Gln	Pro	Ile	Pro	Ala	Ser	Glu	Arg	Arg	Lys	Ile	Ala	
	210					215					220					
Leu	Phe	Thr	Asn	Val	Ala	Glu	Lys	Ala	Val	Ile	Ser	Ala	Ile	Asp	Ala	
225					230					235					240	
Asp	Thr	Ile	Tyr	Arg	Ile	Pro	Leu	Leu	Leu	Arg	Glu	Gln	Gly	Leu	Asp	
				245					250					255		
Asp	Leu	Val	Val	Asp	Gln	Leu	Arg	Leu	Asp	Val	Pro	Ala	Ala	Asp	Leu	
			260					265					270			
Ser	Ala	Trp	Glu	Lys	Val	Val	Asp	Gly	Leu	Thr	His	Pro	Thr	Asp	Glu	
		275					280					285				
Val	Ser	Ile	Ala	Ile	Val	Gly	Lys	Tyr	Val	Asp	His	Thr	Asp	Ala	Tyr	
		290				295					300					
Lys	Ser	Leu	Asn	Glu	Ala	Leu	Ile	His	Ala	Gly	Ile	His	Thr	Arg	His	
305					310					315					320	
Lys	Val	Gln	Ile	Ser	Tyr	Ile	Asp	Ser	Glu	Thr	Ile	Glu	Ala	Glu	Gly	
				325					330					335		
Thr	Ala	Lys	Leu	Lys	Asn	Val	Asp	Ala	Ile	Leu	Val	Pro	Gly	Gly	Phe	
			340					345					350			
Gly	Glu	Arg	Gly	Val	Glu	Gly	Lys	Ile	Ser	Thr	Val	Arg	Phe	Ala	Arg	
		355					360					365				
Glu	Asn	Lys	Ile	Pro	Tyr	Leu	Gly	Ile	Cys	Leu	Gly	Met	Gln	Ser	Ala	
	370					375					380					

Val Ile Glu Phe Ala Arg Asn Val Val Gly Leu Glu Gly Ala His Ser  
 385 390 395 400  
 Thr Glu Phe Leu Pro Lys Ser Pro His Pro Val Ile Gly Leu Ile Thr  
 405 410 415  
 Glu Trp Met Asp Glu Ala Gly Glu Leu Val Thr Arg Asp Glu Asp Ser  
 420 425 430  
 Asp Leu Gly Gly Thr Met Arg Leu Gly Ala Gln Lys Cys Arg Leu Lys  
 435 440 445  
 Ala Asp Ser Leu Ala Phe Gln Leu Tyr Gln Lys Asp Val Ile Thr Glu  
 450 455 460  
 Arg His Arg His Arg Tyr Glu Phe Asn Asn Gln Tyr Leu Lys Gln Leu  
 465 470 475 480  
 Glu Ala Ala Gly Met Lys Phe Ser Gly Lys Ser Leu Asp Gly Arg Leu  
 485 490 495  
 Val Glu Ile Ile Glu Leu Pro Glu His Pro Trp Phe Leu Ala Cys Gln  
 500 505 510  
 Phe His Pro Glu Phe Thr Ser Thr Pro Arg Asn Gly His Ala Leu Phe  
 515 520 525  
 Ser Gly Phe Val Glu Ala Ala Ala Lys His Lys Thr Gln Gly Thr Ala  
 530 535 540

<210> 13

<211> 891

<212> DNA

<213> Methylobionas 16a

<220>

<223> ORF7

<400> 13

```

atgagtaaat tgaaagccta cctgaccgtc tgccaagaac gcgtcgagcg cgcgctggac 60
gccggtctgc ctgccgaaaa catactgcca caaaccttgc atcaggccat gcgctattcc 120
gtattgaacg gcggcaaacg caccgggcc ttgttgactt atgcgaccgg tcaggctttg 180
ggcttgccgg aaaacgtgct ggatgcgccg gcttgcgcg tagaattcat ccatgtgtat 240
tcgctgattc acgacgatct gccggccatg gacaacgatg atctgcgccg cggcaaaccg 300
acctgtcaca aggtttacga cgaggccacc gccatttttg ccggcgacgc actgcaggcg 360
ctggcctttg aagttctggc caacgacccc ggcatcaccg tcgatgcccc ggctcgctg 420
aaaatgatca cggctttgac ccgcgccagc ggctctcaag gcatgggtgg cggtaagcc 480
atcgatctcg gctcgtcgcg ccgcaaattg acgctgccgg aactcgaaaa catgcatatc 540
cacaagactg gcgccctgat ccgcgccagc gtcaatctgg cggcattatc caaacccgat 600
ctggatactt gcgtcgccaa gaaactggat cactatgcca aatgcatagg cttgtcgttc 660
caggtcaaag acgacattct cgacatcgaa gccgacaccg cgacactcgg caagactcag 720
ggcaaggaca tcgataacga caaacccgacc taccttgccg tattgggcat ggctggcgcc 780
aaacaaaaag cccaggaatt gcacgaacaa gcagtcgaaa gcttaacggg atttggcagc 840
gaagccgacc tgctgcgcgga actatcgctt tacatcatcg agcgcacgca c 891
  
```

<210> 14

<211> 297

<212> PRT

<213> Methylomonas 16a

<220>

<223> Amino acid sequences encoded by ORF7

<400> 14

```
Met Ser Lys Leu Lys Ala Tyr Leu Thr Val Cys Gln Glu Arg Val Glu
  1           5           10           15

Arg Ala Leu Asp Ala Arg Leu Pro Ala Glu Asn Ile Leu Pro Gln Thr
      20           25           30

Leu His Gln Ala Met Arg Tyr Ser Val Leu Asn Gly Gly Lys Arg Thr
      35           40           45

Arg Pro Leu Leu Thr Tyr Ala Thr Gly Gln Ala Leu Gly Leu Pro Glu
      50           55           60

Asn Val Leu Asp Ala Pro Ala Cys Ala Val Glu Phe Ile His Val Tyr
      65           70           75           80

Ser Leu Ile His Asp Asp Leu Pro Ala Met Asp Asn Asp Asp Leu Arg
      85           90           95

Arg Gly Lys Pro Thr Cys His Lys Ala Tyr Asp Glu Ala Thr Ala Ile
      100          105          110

Leu Ala Gly Asp Ala Leu Gln Ala Leu Ala Phe Glu Val Leu Ala Asn
      115          120          125

Asp Pro Gly Ile Thr Val Asp Ala Pro Ala Arg Leu Lys Met Ile Thr
      130          135          140

Ala Leu Thr Arg Ala Ser Gly Ser Gln Gly Met Val Gly Gly Gln Ala
      145          150          155          160

Ile Asp Leu Gly Ser Val Gly Arg Lys Leu Thr Leu Pro Glu Leu Glu
      165          170          175

Asn Met His Ile His Lys Thr Gly Ala Leu Ile Arg Ala Ser Val Asn
      180          185          190

Leu Ala Ala Leu Ser Lys Pro Asp Leu Asp Thr Cys Val Ala Lys Lys
      195          200          205

Leu Asp His Tyr Ala Lys Cys Ile Gly Leu Ser Phe Gln Val Lys Asp
      210          215          220

Asp Ile Leu Asp Ile Glu Ala Asp Thr Ala Thr Leu Gly Lys Thr Gln
      225          230          235          240

Gly Lys Asp Ile Asp Asn Asp Lys Pro Thr Tyr Pro Ala Leu Leu Gly
      245          250          255

Met Ala Gly Ala Lys Gln Lys Ala Gln Glu Leu His Glu Gln Ala Val
      260          265          270

Glu Ser Leu Thr Gly Phe Gly Ser Glu Ala Asp Leu Leu Arg Glu Leu
      275          280          285
```

Ser Leu Tyr Ile Ile Glu Arg Thr His  
290 295

<210> 15  
<211> 1533  
<212> DNA  
<213> Methylobionas 16a

<220>  
<223> ORF8

<400> 15  
atggccaaca ccaaacacat catcatcgtc ggcgcgggtc ccggcgggact ttgcgcgggc 60  
atgttgctga gccagcgcg cttcaaggta tcgatttttcg acaaacatgc agaaatcggc 120  
ggcgcgaacc gccgatcaa catgaacggc tttaccttcg ataccgggtcc gacattcttg 180  
ttgatgaaag gcgtgctgga cgaaatgttc gaactgtgcg agcgccgtag cgaggattat 240  
ctggaattcc tgccgctaag cccgatgtac cgcctgctgt acgacgaccg cgacatcttc 300  
gtctattccg accgcgagaa catgcgcgcc gaattgcaac gggatttcga cgaaggcacg 360  
gacggctacg aacagttcat ggaacaggaa cgcaaacgct tcaacgcgct gtatccctgc 420  
atcacccgcg attattccag cctgaaatcc tttttgtcgc tggacttgat caaggccctg 480  
ccgtggctgg cttttccgaa aagcgtgttc aataatctcg gccagtattt caaccaggaa 540  
aaaatgcgcc tggccttttg ctttcagtc cagtatctgg gcatgtcgcc gtgggaatgc 600  
ccggcactgt ttacgatgct gccctatctg gagcacgaat acggcattta tcacgtcaaa 660  
ggcggcctga accgcacgc gcggcgatg gcgcaagtga tcgcggaaaa cggcgccgaa 720  
attcacttga acagcgaaat cgagtcgctg atcatcgaaa acggcgctgc caaggcgctc 780  
aaattacaac atggcgcgga gctgcgcgcc gacgaagtca tcatcaacgc ggattttgcc 840  
cacgcgatga cgcacatggt caaacggggc gtcttgaaaa aatacaccac ggaaaacctg 900  
aagcagcgcg agtattcttg ttgcaccttc atgctgtatc tgggtttgga caagatttac 960  
gatctgccgc accataccat cgtgtttgcc aaggattaca ccaccaatat ccgcaacatt 1020  
ttcgacaaca aaacctgac ggacgatttt tctgtttacg tgcaaaacgc cagcgccagc 1080  
gacgacagcc tagcgccagc cggcaaatcg gcgctgtacg tgctgggtgc gatgcccac 1140  
aacgacagcg gcctggactg gcaggcgcat tgccaaaacg tcgcggaaca ggtgttgga 1200  
acgctgggcg cgcgactggg attgagcgac atcagagccc atatcgaatg cgaaaaaatc 1260  
atcacgccgc aaacctggga aacggacgaa cacgtttaca agggcgccac tttcagtttg 1320  
tcgcacaagt tcagccaaat gctgtactgg cggccgcaca accgtttcga ggaactggcc 1380  
aattgctatc tggtcggcg cggcacgcac cccggtagcg gtttgccgac catctacgaa 1440  
tcggcgcgga tttcgccaa gctgatttcc cagaaacatc gggtagaggt caaggacata 1500  
gcacacagcg cctggctgaa aaaagccaaa gcc 1533

<210> 16  
<211> 511  
<212> PRT  
<213> Methylobionas 16a

<220>  
<223> Amino acid sequences encoded by ORF8

<400> 16  
Met Ala Asn Thr Lys His Ile Ile Ile Val Gly Ala Gly Pro Gly Gly  
1 5 10 15  
Leu Cys Ala Gly Met Leu Leu Ser Gln Arg Gly Phe Lys Val Ser Ile  
20 25 30  
Phe Asp Lys His Ala Glu Ile Gly Gly Arg Asn Arg Pro Ile Asn Met  
35 40 45

Asn	Gly	Phe	Thr	Phe	Asp	Thr	Gly	Pro	Thr	Phe	Leu	Leu	Met	Lys	Gly	50	55	60	
Val	Leu	Asp	Glu	Met	Phe	Glu	Leu	Cys	Glu	Arg	Arg	Ser	Glu	Asp	Tyr	65	70	75	80
Leu	Glu	Phe	Leu	Pro	Leu	Ser	Pro	Met	Tyr	Arg	Leu	Leu	Tyr	Asp	Asp	85	90	95	
Arg	Asp	Ile	Phe	Val	Tyr	Ser	Asp	Arg	Glu	Asn	Met	Arg	Ala	Glu	Leu	100	105	110	
Gln	Arg	Val	Phe	Asp	Glu	Gly	Thr	Asp	Gly	Tyr	Glu	Gln	Phe	Met	Glu	115	120	125	
Gln	Glu	Arg	Lys	Arg	Phe	Asn	Ala	Leu	Tyr	Pro	Cys	Ile	Thr	Arg	Asp	130	135	140	
Tyr	Ser	Ser	Leu	Lys	Ser	Phe	Leu	Ser	Leu	Asp	Leu	Ile	Lys	Ala	Leu	145	150	155	160
Pro	Trp	Leu	Ala	Phe	Pro	Lys	Ser	Val	Phe	Asn	Asn	Leu	Gly	Gln	Tyr	165	170	175	
Phe	Asn	Gln	Glu	Lys	Met	Arg	Leu	Ala	Phe	Cys	Phe	Gln	Ser	Lys	Tyr	180	185	190	
Leu	Gly	Met	Ser	Pro	Trp	Glu	Cys	Pro	Ala	Leu	Phe	Thr	Met	Leu	Pro	195	200	205	
Tyr	Leu	Glu	His	Glu	Tyr	Gly	Ile	Tyr	His	Val	Lys	Gly	Gly	Leu	Asn	210	215	220	
Arg	Ile	Ala	Ala	Ala	Met	Ala	Gln	Val	Ile	Ala	Glu	Asn	Gly	Gly	Glu	225	230	235	240
Ile	His	Leu	Asn	Ser	Glu	Ile	Glu	Ser	Leu	Ile	Ile	Glu	Asn	Gly	Ala	245	250	255	
Ala	Lys	Gly	Val	Lys	Leu	Gln	His	Gly	Ala	Glu	Leu	Arg	Gly	Asp	Glu	260	265	270	
Val	Ile	Ile	Asn	Ala	Asp	Phe	Ala	His	Ala	Met	Thr	His	Leu	Val	Lys	275	280	285	
Pro	Gly	Val	Leu	Lys	Lys	Tyr	Thr	Pro	Glu	Asn	Leu	Lys	Gln	Arg	Glu	290	295	300	
Tyr	Ser	Cys	Ser	Thr	Phe	Met	Leu	Tyr	Leu	Gly	Leu	Asp	Lys	Ile	Tyr	305	310	315	320
Asp	Leu	Pro	His	His	Thr	Ile	Val	Phe	Ala	Lys	Asp	Tyr	Thr	Thr	Asn	325	330	335	
Ile	Arg	Asn	Ile	Phe	Asp	Asn	Lys	Thr	Leu	Thr	Asp	Asp	Phe	Ser	Phe	340	345	350	
Tyr	Val	Gln	Asn	Ala	Ser	Ala	Ser	Asp	Asp	Ser	Leu	Ala	Pro	Ala	Gly	355	360	365	

Lys Ser Ala Leu Tyr Val Leu Val Pro Met Pro Asn Asn Asp Ser Gly  
 370 375 380  
 Leu Asp Trp Gln Ala His Cys Gln Asn Val Arg Glu Gln Val Leu Asp  
 385 390 395 400  
 Thr Leu Gly Ala Arg Leu Gly Leu Ser Asp Ile Arg Ala His Ile Glu  
 405 410 415  
 Cys Glu Lys Ile Ile Thr Pro Gln Thr Trp Glu Thr Asp Glu His Val  
 420 425 430  
 Tyr Lys Gly Ala Thr Phe Ser Leu Ser His Lys Phe Ser Gln Met Leu  
 435 440 445  
 Tyr Trp Arg Pro His Asn Arg Phe Glu Glu Leu Ala Asn Cys Tyr Leu  
 450 455 460  
 Val Gly Gly Gly Thr His Pro Gly Ser Gly Leu Pro Thr Ile Tyr Glu  
 465 470 475 480  
 Ser Ala Arg Ile Ser Ala Lys Leu Ile Ser Gln Lys His Arg Val Arg  
 485 490 495  
 Phe Lys Asp Ile Ala His Ser Ala Trp Leu Lys Lys Ala Lys Ala  
 500 505 510

<210> 17  
 <211> 1491  
 <212> DNA  
 <213> Methylomonas 16a

<220>  
 <223> ORF9

<400> 17  
 atgaactcaa atgacaacca acgcgtgacg gtgatcggcg ccggcctcgg cggcctgtcc 60  
 gccgctatct cgctggccac ggccggcctt tccgtgcaac tcatcgaaaa aaacgacaag 120  
 gtcggcggca agctcaacat catgacccaa gacggcttta ccttcgatct ggggcccgtc 180  
 attttgacga tgccgcacat ctttgaggcc ttgttcacag gggccggcaa aaacatggcc 240  
 gattacgtgc aaatccagaa agtcgaaccg cactggcgca atttcttcga ggacggtagc 300  
 gtgatcgact tgtgcaaga cgccgaaacc cagcgccgag agctggataa acttggcccc 360  
 ggcacttacg cgcaattcca gcgctttctg gactattcga aaaacctctg cacggaaaacc 420  
 gaagccgggt acttcgcca gggcctggac ggcttttggg atttactcaa gttttacggc 480  
 ccgctccgca gcctgctgag tttcgacgtc ttccgcagca tggaccaggg cgtgcgccgc 540  
 tttatttccg atcccaagtt ggtcgaaatc ctgaattact tcatcaaata cgtcggctcc 600  
 tcgccttacg atgcgcccgc cttgatgaac ctgctgcctt acattcaata tcattacggc 660  
 ctgtggtacg tgaaaggcgg catgtatggc atggcgaggg ccatggaaaa actggccgtg 720  
 gaattgggag tcgagattcg tttgatgacc gaggtgtcgg aaatccaaaa acaggacggc 780  
 agagcctgag ccgtaaaagt ggcgaacggc gacgtgctgc cggccgacat cgtgggtgtc 840  
 aacatggaag tgattccggc gatggaaaaa ctgctgcgca gcccgccag cgaactgaaa 900  
 aaaatgcagc gcttcgagcc tagctgttcc ggctgggtgc tgcacttggg cgtggacagg 960  
 ctgtatccgc aactggcgca ccacaatttc ttttattccg atcatccgag cgaacatttc 1020  
 gatgcggtat tcaaaagcca tcgcctgtcg gacgatccga ccatttatct ggtcgcgccg 1080  
 tgcaagaccg accccgcccc ggccggcgcc ggctgcgaga tcatcaaaat cctgccccat 1140  
 atcccgcacc tcgaccccgca caaactgctg accgccgagg attattcagc cttgcgcgag 1200  
 cgggtgctgg tcaaaactcg acgcatgggc ctgacggatt tacgccaaca catcgtgacc 1260  
 gaagaatact ggacgccgct ggatattcag gccaaatatt attcaaacca gggctcgtat 1320  
 tacggcgtgg tcgccgaccg cttcaaaaac ctgggtttca aggcacctca acgcagcagc 1380



gaattatcca atctgtatatt cgtcggcggc agcgtcaatc ccggcggcgg catgccgatg 1440  
 gtgacgctgt ccgggcaatt ggtgaggac aagattgtgg cggatttgca a 1491

<210> 18  
 <211> 497  
 <212> PRT  
 <213> Methylomonas 16a

<220>  
 <223> Amino acid sequences encoded by ORF9

<400> 18  
 Met Asn Ser Asn Asp Asn Gln Arg Val Ile Val Ile Gly Ala Gly Leu  
   1                  5                  10                  15  
 Gly Gly Leu Ser Ala Ala Ile Ser Leu Ala Thr Ala Gly Phe Ser Val  
                   20                  25                  30  
 Gln Leu Ile Glu Lys Asn Asp Lys Val Gly Gly Lys Leu Asn Ile Met  
           35                  40                  45  
 Thr Lys Asp Gly Phe Thr Phe Asp Leu Gly Pro Ser Ile Leu Thr Met  
   50                  55                  60  
 Pro His Ile Phe Glu Ala Leu Phe Thr Gly Ala Gly Lys Asn Met Ala  
   65                  70                  75                  80  
 Asp Tyr Val Gln Ile Gln Lys Val Glu Pro His Trp Arg Asn Phe Phe  
                   85                  90                  95  
 Glu Asp Gly Ser Val Ile Asp Leu Cys Glu Asp Ala Glu Thr Gln Arg  
           100                  105                  110  
 Arg Glu Leu Asp Lys Leu Gly Pro Gly Thr Tyr Ala Gln Phe Gln Arg  
   115                  120                  125  
 Phe Leu Asp Tyr Ser Lys Asn Leu Cys Thr Glu Thr Glu Ala Gly Tyr  
   130                  135                  140  
 Phe Ala Lys Gly Leu Asp Gly Phe Trp Asp Leu Leu Lys Phe Tyr Gly  
   145                  150                  155                  160  
 Pro Leu Arg Ser Leu Leu Ser Phe Asp Val Phe Arg Ser Met Asp Gln  
           165                  170                  175  
 Gly Val Arg Arg Phe Ile Ser Asp Pro Lys Leu Val Glu Ile Leu Asn  
           180                  185                  190  
 Tyr Phe Ile Lys Tyr Val Gly Ser Ser Pro Tyr Asp Ala Pro Ala Leu  
   195                  200                  205  
 Met Asn Leu Leu Pro Tyr Ile Gln Tyr His Tyr Gly Leu Trp Tyr Val  
   210                  215                  220  
 Lys Gly Gly Met Tyr Gly Met Ala Gln Ala Met Glu Lys Leu Ala Val  
   225                  230                  235                  240  
 Glu Leu Gly Val Glu Ile Arg Leu Asp Ala Glu Val Ser Glu Ile Gln  
           245                  250                  255

Lys Gln Asp Gly Arg Ala Cys Ala Val Lys Leu Ala Asn Gly Asp Val  
 260 265 270  
 Leu Pro Ala Asp Ile Val Val Ser Asn Met Glu Val Ile Pro Ala Met  
 275 280 285  
 Glu Lys Leu Leu Arg Ser Pro Ala Ser Glu Leu Lys Lys Met Gln Arg  
 290 295 300  
 Phe Glu Pro Ser Cys Ser Gly Leu Val Leu His Leu Gly Val Asp Arg  
 305 310 315 320  
 Leu Tyr Pro Gln Leu Ala His His Asn Phe Phe Tyr Ser Asp His Pro  
 325 330 335  
 Arg Glu His Phe Asp Ala Val Phe Lys Ser His Arg Leu Ser Asp Asp  
 340 345 350  
 Pro Thr Ile Tyr Leu Val Ala Pro Cys Lys Thr Asp Pro Ala Gln Ala  
 355 360 365  
 Pro Ala Gly Cys Glu Ile Ile Lys Ile Leu Pro His Ile Pro His Leu  
 370 375 380  
 Asp Pro Asp Lys Leu Leu Thr Ala Glu Asp Tyr Ser Ala Leu Arg Glu  
 385 390 395 400  
 Arg Val Leu Val Lys Leu Glu Arg Met Gly Leu Thr Asp Leu Arg Gln  
 405 410 415  
 His Ile Val Thr Glu Glu Tyr Trp Thr Pro Leu Asp Ile Gln Ala Lys  
 420 425 430  
 Tyr Tyr Ser Asn Gln Gly Ser Ile Tyr Gly Val Val Ala Asp Arg Phe  
 435 440 445  
 Lys Asn Leu Gly Phe Lys Ala Pro Gln Arg Ser Ser Glu Leu Ser Asn  
 450 455 460  
 Leu Tyr Phe Val Gly Gly Ser Val Asn Pro Gly Gly Gly Met Pro Met  
 465 470 475 480  
 Val Thr Leu Ser Gly Gln Leu Val Arg Asp Lys Ile Val Ala Asp Leu  
 485 490 495

Gln

<210> 19

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence:primer

<400> 19

aaggatccgc gtattcgtag tc

22

<210> 20  
 <211> 40  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence:primer  
  
 <400> 20  
 ctggatccga tctagaaata ggctcgagtt gtcgttcagg 40  
  
  
 <210> 21  
 <211> 30  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence:primer  
  
 <400> 21  
 aaggatccta ctcgagctga catcagtgc 30  
  
  
 <210> 22  
 <211> 22  
 <212> DNA  
 <213> Artificial Sequence  
  
 <220>  
 <223> Description of Artificial Sequence:primer  
  
 <400> 22  
 gctctagatg caaccagaat cg 22  
  
  
 <210> 23  
 <211> 954  
 <212> DNA  
 <213> Methylomonas 16a  
  
 <400> 23  
 atgcaaatacg tactcgcaaa cccccgtgga ttctgtgccg gcgtggaccg ggccattgaa 60  
 attgtcgatc aagccatcga agccttttgt gcgccgattt atgtgcggca cgaggtggtg 120  
 cataaccgca ccgtgggtcga tggactgaaa caaaaagggtg cgggtgttcac cgaggaacta 180  
 agcgatgtgc cgggtgggttc ctacttgatt ttcagcgcgc acggcgtatc caaggagggtg 240  
 caacaggaag ccgaggagcg ccagttgacg gtattcgatg cgacttgacc gctggtgacc 300  
 aaagtgcaca tgcaggttgc caagcatgcc aaacagggcc gagaagtgat tttgatcggc 360  
 cacgccggtc atccggaagt ggaaggcacg atgggccagt atgaaaaatg caccgaaggc 420  
 ggcggcattt atctggtcga aactccggaa gacgtacgca atttgaaagt caacaatccc 480  
 aatgatctgg cctatgtgac gcagacgacc ttgtcgatga ccgacaccaa ggtcatggtg 540  
 gatgcgttac gcgaacaatt tccgtccatt aaggagcaaa aaaaggacga tatttggtac 600  
 gcgacgcaaa accgtcagga tgcggtgcat gatctggcca agatttccga cctgattctg 660  
 gttgtcggct ctcccaatag ttcgaattcc aaccgtttgc gtgaaatcgc cgtgcaactc 720  
 ggtaaaccgc cttatattgat cgatacttac caggatttga agcaagattg gctggaggga 780  
 attgaagtag tcgggggttac cgcgggcgct tcggcgcccg aagtgttggt gcaggaagtg 840  
 atcgatcaac tgaaggcatg gggcggcgaa accacttcgg tcagagaaaa cagcggcatc 900  
 gaggaaaagg tagtcttttc gattcccaag gagttgaaaa aacatatgca agcg 954

<210> 24  
 <211> 318  
 <212> PRT  
 <213> Methylobionas 16a

<400> 24

Met	Gln	Ile	Val	Leu	Ala	Asn	Pro	Arg	Gly	Phe	Cys	Ala	Gly	Val	Asp	1	5	10	15
Arg	Ala	Ile	Glu	Ile	Val	Asp	Gln	Ala	Ile	Glu	Ala	Phe	Gly	Ala	Pro	20	25	30	
Ile	Tyr	Val	Arg	His	Glu	Val	Val	His	Asn	Arg	Thr	Val	Val	Asp	Gly	35	40	45	
Leu	Lys	Gln	Lys	Gly	Ala	Val	Phe	Ile	Glu	Glu	Leu	Ser	Asp	Val	Pro	50	55	60	
Val	Gly	Ser	Tyr	Leu	Ile	Phe	Ser	Ala	His	Gly	Val	Ser	Lys	Glu	Val	65	70	75	80
Gln	Gln	Glu	Ala	Glu	Glu	Arg	Gln	Leu	Thr	Val	Phe	Asp	Ala	Thr	Cys	85	90	95	
Pro	Leu	Val	Thr	Lys	Val	His	Met	Gln	Val	Ala	Lys	His	Ala	Lys	Gln	100	105	110	
Gly	Arg	Glu	Val	Ile	Leu	Ile	Gly	His	Ala	Gly	His	Pro	Glu	Val	Glu	115	120	125	
Gly	Thr	Met	Gly	Gln	Tyr	Glu	Lys	Cys	Thr	Glu	Gly	Gly	Gly	Ile	Tyr	130	135	140	
Leu	Val	Glu	Thr	Pro	Glu	Asp	Val	Arg	Asn	Leu	Lys	Val	Asn	Asn	Pro	145	150	155	160
Asn	Asp	Leu	Ala	Tyr	Val	Thr	Gln	Thr	Thr	Leu	Ser	Met	Thr	Asp	Thr	165	170	175	
Lys	Val	Met	Val	Asp	Ala	Leu	Arg	Glu	Gln	Phe	Pro	Ser	Ile	Lys	Glu	180	185	190	
Gln	Lys	Lys	Asp	Asp	Ile	Cys	Tyr	Ala	Thr	Gln	Asn	Arg	Gln	Asp	Ala	195	200	205	
Val	His	Asp	Leu	Ala	Lys	Ile	Ser	Asp	Leu	Ile	Leu	Val	Val	Gly	Ser	210	215	220	
Pro	Asn	Ser	Ser	Asn	Ser	Asn	Arg	Leu	Arg	Glu	Ile	Ala	Val	Gln	Leu	225	230	235	240
Gly	Lys	Pro	Ala	Tyr	Leu	Ile	Asp	Thr	Tyr	Gln	Asp	Leu	Lys	Gln	Asp	245	250	255	
Trp	Leu	Glu	Gly	Ile	Glu	Val	Val	Gly	Val	Thr	Ala	Gly	Ala	Ser	Ala	260	265	270	

Pro	Glu	Val	Leu	Val	Gln	Glu	Val	Ile	Asp	Gln	Leu	Lys	Ala	Trp	Gly
		275					280					285			
Gly	Glu	Thr	Thr	Ser	Val	Arg	Glu	Asn	Ser	Gly	Ile	Glu	Glu	Lys	Val
	290					295					300				
Val	Phe	Ser	Ile	Pro	Lys	Glu	Leu	Lys	Lys	His	Met	Gln	Ala		
305					310					315					